数理解析研究所講究録
第 699 巻 1989 年 132-144

ニューラルネットワークの数理的性質
Mathematical Properties of Neural Networks

麻生　英樹
Hideki Asoh (asoh@etl.JUNET)
電子技術総合研究所
Electrotechnical Laboratory

Abstcact

現在までに提案されているニューラルネットワークによる情報処理および情報処理の学習のモデルの中から代表的と思われるものを選び，それらの数理的な性質についての結果を紹介する．

1　はじめに

生体の神経系を数式あるいは，計算機プログラムなどによってモデル化するという試みは多岐にわたっている．そうしたバラエティの由来の一つは，モデルを作るにあたっての目的意識の差異である．たとえば，実際の神経細胞の動作を解明することを目的として，単一の神経細胞の一部の性質を詳細にモデル化することが行われる．一方，かなり大規模な神経ネットワークの動作を解明することを目的とする場合には，個々の神経細胞（計算ユニット）にはかなり簡単なモデルが使われることが多い．

本稿では,「比較的大規模な神経ネットワークによる情報処理のメカニズムを解明する」という目的意識のもとに提案され，現在よく取り上げられているいくつかの数理モデルの数理的な性質を解説する．従って，生理的なデータとの一致を目的とするモデルや，ニューラルネットワークの非線形力学系としてのダイナミクスの解析のみを目的とするモデルなどはここでの視野外となる．

なお，こうした研究の分野を指示する言葉として，ニューラルネット，ニューロコンピューティングなどが，研究者あるいは文脈によってかなり自由に使われている．以下では，生体の神経系を，情報処理の研究という特定の目的の下に，簡単にモデル化したものをニューラルネット（ワーク）と呼ぶ．

本稿は，人工知能学会誌に投稿中の解説に加筆補正を加えたものである．

2　ニューラルネットによる情報処理

神経系による情報処理のメカニズムを理解するためのモデルには，神経系のモデルとともに，その働き方のモデル，すなわち，そのモデルによる情報処理のモデルが含まれる．この情報処理のモデル化という問題自体は,「情報処理とは何であるか」というような深い内容を持つものであるが，ここではそちらには踏み込むことはせずに現状を紹介するにとどめる．

解説の便宜のために，現在までに提案されている主な情報処理のモデルの中から代表的なものとして次の四つを選んだ．

1. 階層的なネットワークによる入出力パターン変換のモデル

2. 回帰的な結合を含むネットワークによるパターンの時系列生成，処理のモデル

3. 相互結合ネットワークによる連想記憶のモデル

4. 相互結合ネットワークによる最適解探索のモデル

このうち，1）と2）は近く，3）と4）も近い．以下，それぞれのモデルについて，現在までに知られている数理的性質を示す．

2.1　階層的ネットワークによる入出力パターン変換

このモデルは，パーセプトロン[29] [30] に代表されるものである．ニューラルネットとしては，入力層，出力層と複数の中間層から成る多層の階層的な構造のものを用いる．このようなネットワークでは，入力層に提示される入力パターンは層を経るごとに変換され出力層に到る．この変換をこのネットワークの行う情報処理と考えるわけである．入力層を除く各層のユニット（神経細胞）とそれらの相互作用（シナプス）のモデルとしては，次のような簡単なものが使われる（入力層のユニットは，通常，入力パターンをそのまま出力するバッファのようなものである）．

$$v_j = f(u_j), \tag{1}$$

$$u_j = \sum_i w_{ij} v_j - \theta_j \tag{2}$$

ここで，v_jはユニット j の出力，w_{ij} はユニット i から j への結合の重みと呼ばれる実数値である．θ_jはユニット j をしきい値，u_jをユニット jへの入力の総和などと呼ぶ．fはユニットの入出力関数，特性関数などと呼ばれる関数で，通常，生理学の知見および数学的な扱い易さなどの理由で，ロジスティック関数 $f(x) = 1/(1+e^{-x})$ のようなシグモイド（S字形）の，単調増加で飽和特性

を持つ関数が使われる．こうしたユニットは，式（２）の線形性と，式（１）の f の非線形性とをあわせ持つため，準線形のユニットと呼ばれる [32]．

このモデルにおいてまず問題になるのは，入出力変換の能力とネットワークの構造との関係である．出力層のユニットの間には相互作用がないので，議論を単純化するために出力層のユニットが一つの場合を考えればよい．入力層のユニットの数を n とすれば，ネットワークは n 変数の関数を計算していることになる．したがって問題はこうである：

- どのような規模／構造のネットワークによって，どのような関数（対応づけ）が実現できるのか？

たとえば，ユニットの入出力関数としてしきい関数（入力の総和が負なら値０，正なら値１をとる）を用いたネットワークの場合には，中間層が無ければ，入力パターンの空間（ n 次元）をある超平面で切り，その一方に含まれる入力に対して１を他方に含まれる入力に対して０を出力するような関数関係しか実現できないことが簡単にわかる（入力の総和を求める計算が入力値に対して線形なためである）．この条件は**線形分離性の条件**として知られている．しかし，一つのユニットによってそのユニットへの入力全体のＡＮＤ／ＯＲを計算することができるため，十分にたくさん（たとえば 2^n 個）のユニットを持つ中間層を一層だけ用意すれば，任意の n 変数論理関数 $g:\ \{0,1\}^n \to \{0,1\}$ を実現できることも簡単にわかる．また，Minsky と Papert らは，主に層の間の結合に制約がある場合の能力について詳しい考察を行い，ある種の制約（有限次数）のもとでは実現できない関数があることなどを数理的に示している [25]．

最近の最も明快な結果は次の定理であろう [8][15]

定理１ シグモイド状の入出力関数を持つ準線形のユニットを使った中間層を一層持つ階層的なネットワークによって，任意の連続関数を任意の精度で近似実現することができる．

この定理が述べていることは，どのような関数（入出力変換）であっても，十分にたくさんのユニットを持つ中間層を一層用意すれば，望むだけの精度でそれを近似するような結合の重みの決め方が存在する，ということである．この定理の証明はある程度構成的であるが，それが与える方法は多くの場合に非常にたくさんの中間層ユニットを必要とする．

もちろん，以上のような結果は上の問題に対するそれぞれの答え方であり，これらによって問題が完全に解決しているわけではまったくない．たとえば，ある関数を十分な精度で近似するための最小の構成，後で述べる変換の学習法との関連において，より簡単に学習できるためのネットワークの構造など多くの問題が未解決である．　実際の神経系との関連を気にしないならば，たとえば，中間層のユニットの入出力関数を三角関数にとり，最終層のユニットの入出力関数を恒

等関数にとることもできる．この場合には，三層のネットワークで関数を近似することは，その関数をフーリエ展開することに対応する．したがって，ネットワークによる関数近似の性質ははっきりとしている．

2.2 回帰的な結合を含むネットワークによるパターンの時系列の生成，処理

このモデルは，上のモデルを時間方向に拡張したものである．従って，これに関する問題も基本的には同じで，

- どのような規模／構造のネットワークによって，どのような時系列が実現できるのか？
- どのような規模／構造のネットワークによって，どのような時系列処理（時系列の対応づけ）が実現できるか？

というものである．

原理的な能力に関しては次の古い定理がある．

定理 2 ニューラルネットによって任意の有限オートマトンをシミュレートすることができる [19]．

また，定理 1 の結果を利用すれば，入力から出力を計算するのに適当な時間遅れΔt を伴うような階層的なネットワークの出力を入力にフィードバックさせることで，任意のシステム

$$\mathbf{x}(t+\Delta t) = f(\mathbf{x}(t) + \mathbf{s}(t)) \tag{3}$$

を実現することができるのは明らかである．ここで，$\mathbf{x}(t)$ と $\mathbf{s}(t)$ は，それぞれ，時刻 t におけるネットワークの出力とネットワークへの外部からの入力である．実際的なネットワークの能力ついては，さまざまな構造のネットワークによる数値実験が行われている段階である．

2.3 相互結合ネットワークによる連想記憶のモデル

人間の記憶は，従来の計算機の記憶のような番地によるアクセスではなく，類似した事柄や関連する事柄から記憶内容が想起される,「連想的な記憶」であるといわれている．人間の情報処理について考える上でも，その記憶，すなわち情報の保存と利用の形態が，このような連想的なものであるという点は重要な鍵になると思われる．ニューラルネットによってこうした連想的な記憶のモデルを作るという試みも古くからあり，代表的な研究としては，中野によるアソシアトロン [26]，Kohonen による連想記憶 [20] [21] などがあげられる．また最近では，Hopfield による連想記憶のモデルと解析が提案され [13]，今回のブームの一つの要因となった．

現在までにさまざまな連想記憶のモデルが提案されているが，連想記憶のモデルに対する基本的な問題は，

- 記銘の問題：記憶させたいパターンをどのようにして記憶させるのか？
- 容量の問題：どれくらいの規模のネットワークでいくつくらいのパターンを記憶できるのか？
- 耐雑音，変形の問題：鍵パターンがどの程度まで雑音で汚されたり，変形されたりしても，想起することができるか？

であろう．

こうした性質を数理的に議論するにあたっては，連想記憶のモデルをいくつかの軸によって分類しておくことが有効であり，通常二つの軸がよく使われる．一つ目の軸は，連想記憶を機能による分類で，記憶内容そのものを鍵として想起を行う「自己想起型の連想記憶」と記憶内容と関連づけられている情報を鍵として想起を行う「相互想起型の連想記憶」に分けるものである．もう一つの軸は，連想記憶の実現に使われているニューラルネットの形態によって分けるものである．以下では，相互結合のあるネットワークによる自己想起型の連想記憶に対する上の問題への考察の一部を典型的な例によって紹介する．

n 個のユニットから成るネットワークを考える．ユニットは準線形とし，入出力関数 f としては，二値 $\{-1,+1\}$ をとるしきい関数を使う．すなわち，ユニットへの入力の重みつき総和が正ならば出力 1，負ならば出力は -1 となる．ネットワークは相互結合型，すなわち，任意の二つのユニットの間に結合があるとする．各ユニットの状態変化のタイミングとしては，すべてのユニットが一斉に状態変化する同期型と，各時刻に一つのユニットだけが状態変化する非同期型があるが，それぞれの方法の数理的性質はあまり大きくは違わない．

このようなネットワークの動作は，結合の重み $\{w_{ij}\}$ の値と時刻 0 での各ユニットの出力値（初期状態）$v_i(0)$ によって定まる．こうしたネットワークを自己連想型の連想記憶に使う方法としては，記憶させたいパターンがネットワークの安定な平衡状態となるように結合の重み $\{w_{ij}\}$ を設定し，安定な平衡状態の近くの状態を初期値としてネットワークを動作させた場合に，最終的には状態は安定な平衡状態へ引き込まれるという性質を利用して雑音や変形に耐性のある想起を実現するというものが提案されている．

こうしたモデルにおける記銘の問題とは，記憶させたいパターンが与えられたときに，結合の重み $\{w_{ij}\}$ をどう決めるかという問題である．現在までの研究においてもっともよく用いられている方法は，w_{ij} として記憶させたいパターンの集合の自己相関係数を用いるという方法である [13][20][26]．すなわち，記銘させたい m 個のパターン（n 次元ベクトル）の集合を $\{\mathbf{x}^p\}_{p=1}^{m}$ とするとき，

$$w_{ij} = \sum_{p=1}^{m} \mathbf{x}^p{}_i \mathbf{x}^p{}_j \tag{4}$$

のように結合を定める.

この記銘法について記憶容量や想起能力を評価するために，パターン $\{\mathbf{x}_p\}$ の各要素が独立に，確率 1/2 で値 1 か -1 をとるようにランダムに選ばれた場合についての考察が，神経統計力学，情報理論，スピングラスの統計力学などの立場からなされている [4][5][6][24][35]．それによると，パターンの数 m がユニットの数 n に対して $n/4\log n$ 以下ならば，n を大きくしたときに，すべてのパターンが安定な平衡点になり，ハミング距離で $n/2$ 以下離れたパターンから正しく想起される確率が 1 に近づく [24]，m と n の比が 0.14 を越えると，全てのパターンを安定な平衡点にすることはむずかしい [6]，などの結果が得られている．また，パターンが確率的に独立ではない場合にはさらに容量が増えるという結果もある [37]．

別の記銘法としては，銘記させたいパターン $\{\mathbf{x}^p\}$ を並べた行列を $X = (\mathbf{x}^1 \ldots \mathbf{x}^m)$ とするときに，w_{ij} を

$$w_{ij} = [X(X^t X)^{-1} X^t]_{ij} \tag{5}$$

のように決める方法がある [7][21]．この方法によれば，最大，ユニットの数と同じ数の一次独立なパターンを安定な平衡点にすることができる．また，この時の想起能力の解析も行われている [7]．

2.4 相互結合型ネットワークによる最適解探索

Hopfield は，対称な結合を持つネットワークの動作を，ネットワークの状態から値の定まるある関数 Eを減少させる，という形で特徴づけることができることを示し，この関数 Eを力学の用語を借りて，ネットワークのエネルギーと呼んだ．従って，こうしたネットワークが状態変化するたびに，エネルギー Eの値は減少し，いずれは E の極小を与える状態に到って，ネットワークの状態は変わらなくなる．この性質を利用して，エネルギーがある最適化問題の評価関数に対応するようなネットワークをつくって動作させることによって近似的に最適化問題を解くというアイデアが提案され，いくつかの問題でかなりよい解が得られることが示された [14]．

しかし，通常の状態変化規則に従うネットワークは，エネルギーを単調に減らすため，ネットワークの状態はエネルギーを極小にするものに収束し（一般には極小を与える状態は複数あるため，どこに収束するかは初期状態に依存する．これが前節の連想記憶の原理であった.），必ずしも最適な解を得ることができない．そこで，この極小への収束を避けるための工夫として，

1. よい初期状態をさがす

2. 問題の定式化を工夫して，できるだけ最適解の近くに極小解がないようにする

3. ユニットの動作に確率を導入し，いくぶんかの確率でエネルギーを増加させるような状態変化を許すことにする

などの方法が提案されている．

1）に関して，状態空間の中心付近を初期状態にするとよいという予想がなされた [36]．これは実際にかなりよい結果を与えるが，反例あり，確実なものではない．
2）に関しては，ネットワーク上で数を表現する方法を変えて結果を比較してる例がある [34]．
3）の例としてはボルツマンマシン [12] がよく知られている．そこでは各ユニットが次のような確率に従って次の時刻の出力を確率的に決定する．

$$\text{確率 } p = \frac{1}{1+\exp(-u_i/T)} \quad \text{で出力 } v_i = 1 \tag{6}$$

ここで，u_i は式 (2) で与えられるユニットへの入力の総和，T は動作のランダムさを決める正の実数であり，ネットワークの温度と呼ばれる．

このような，ネットワークの状態遷移はマルコフ的な確率過程に従うことになり，初期状態から十分長い時間が経った後のネットワークの状態を観測したときにある状態 $\mathbf{v}$ が得られる確率の分布 $P(\mathbf{v})$（平衡分布）は

$$E(\mathbf{v}) = -\frac{1}{2}\sum_{ij} w_i j v_i v_j + \sum_i \theta_i v_i \tag{7}$$

として，

$$P(\mathbf{v}) \propto \exp(-E(\mathbf{v})/T) \tag{8}$$

となる．$E(\mathbf{v})$ がネットワークのエネルギーと呼ばれる量である．このような，系のエネルギーが指数関数の肩に乗った形の分布を統計力学では ボルツマン分布と呼ぶ．

この分布の形から，一般に，エネルギーが小さな状態ほど高い確率で観測されることがわかる．さらに，温度が十分に低いときには，エネルギーを最小にする状態が観測される確率が 1 に近いことが期待される．しかし，実際には，温度が低いということは動作が決定的なものに近づくということであり，エネルギー極小の状態に長時間とどまるということでもある（しかし，相対的に，エネルギー最小の状態に滞在する確率が増加する）．そこで，効率よく低い温度での平衡分布を得るために，ネットワークの温度をはじめ高い温度に設定し，徐々に減らしながら状態変化を行わせる方法が提案されている [18]．この方法は，規則正しい結晶構造を得るための手法である焼きなましを模擬しているものであるため，一般に模擬焼きなまし，模擬徐冷などと呼ばれている．

このときに問題になるのは温度を変化させる速度である．理論的に最小値を与える状態への収束確率を 1 にするための十分条件としては，

$$T(t) > \frac{T(0)}{\log(t)} \tag{9}$$

というものが知られている．ここで，t は，ネットワークの全ユニットが平均的に1回状態変化をするのにかかる時間である．

これは，実際的にはかなり遅いため，より速い温度変化スケジュールで十分な方法がいくつか提案されている．例えば，ボルツマンマシンの状態変化規則では，各時刻に状態変化をするユニットは一つであったが，これを複数にして比較的遠くの状態へのジャンプを許すことによって焼きなましのスケジュールを加速するというような提案がある．具体的には，現在の状態を中心とするN次元の Cauchy 分布に従って次の状態の候補を選び，その候補とのエネルギー差によって状態遷移を行うか否かを確率的に決めるような場合には（コーシーマシン）$T(t) \quad > T(0)/(1+t)$ というスケジュールで十分であるという結果がある [33]．また，状態が多値の場合や，より一般の平衡分布を実現するような状態変化規則への拡張も提案されているが [10][22]，こうした拡張や変形によって，いわゆるニューラルネットとの関連は薄れる．

以上，代表的と思われるニューラルネットワークによる情報処理モデルを紹介し，その数理的な性質として知られている結果を述べた．この他にも，たとえば，カオスのフラクタル次元の切り替わりを利用した連想記憶など興味深いモデルが多数提案されているが，その性質や能力はあまり明らかになっていない．

3　ニューラルネットによる学習

前節で紹介したようなさまざまな情報処理のモデルにおいては，個々の情報処理課題を実行するために必要な情報は，最終的にはユニット間の結合の重みとして表現されている．このような情報の源泉としては，

- あらかじめ明示的にそれぞれの結合の重みが与えられる（生得的な情報），
- 課題を実行しながら学習的に重みを調節する（学習によって獲得される情報）

の二つが考えられる．この二つの方法をバランスよく用いることが，柔軟で適応的な情報処理のために必要である．

ニューラルネットワークによるこうした学習のモデルもさまざまなものが提案されている．ある学習方式に対して，数理的に問題になることは，

1. 学習の収束（学習可能な問題のクラス）
2. 収束速度の問題
3. 学習結果の評価の問題

4. 学習結果の一般化の問題

5. 学習法の一般化の問題

などである．以下では，最近最もよくとりあげられている誤差逆伝播学習の数理的な性質に関する結果を簡単に紹介する．

誤差逆伝播学習についてはすでに多くの紹介があるため，ここではアルゴリズムの詳細には触れないが，おおまかに言うと，階層的なネットワークによる対応づけを例から学習するための方法として提案された学習法である [31][32]．この方法は数理的には確率的な最急降下法であり，個々の正解例が与えられるごとに，その正解例に対する誤差を最も減らすような方向に結合の重みを微小量だけ変化させることによって，平均的に，学習用の例全体に対する対応づけの誤差を減らしてゆくというものである [2]．

この学習の収束に関しては，次の定理がある [2]．

定理 3 各回の修正量が十分に小さいならば，学習は各正解例に対する二乗誤差の総和を極小にするような重みに収束する．

このように，学習アルゴリズムは局所最適な重みへの収束しか保証していない．実際に，重みの初期値を非常に特殊な（偏った）値にセットして学習を行わせるとしばしば容易に極小解にトラップされ，最適な解には到らない．

この極小解への収束を避ける方法としては各回の修正量を学習のはじめでは大きくとり，学習が進むとともに小さくしてゆく，という焼きなまし的な方法が考えられるが，これに関する定量的な議論はまだないようである．また，おおまかな正解がわかっている場合には，結合の重みの初期状態をその解に設定することによって，極小解を避けることができる．

学習の速度の評価は中間層の数，対応課題の性質などに依存するため，はっきりとした評価はなされていない．学習を加速する方法はいくつか提案されている [2][11][17][28]．学習結果の評価としては，誤差の二乗和極小という基準がおおよその評価を与えている．このことから，各ユニットが入出力関数として恒等関数を用いている場合（線形のユニット）には，誤差逆伝播学習によって生成される中間層上の入力パターンの表現と，従来からある線形の多変量解析手法との間に密接な対応がつくことが知られている [9]．

学習結果の一般化の能力，すなわち，少数の例から一般的な規則を学習する能力についても，いろいろな研究が行われているが，定量的な結果はまだほとんどない．中間層のユニットの数や層の数を必要以上に増やすことは過剰適応を招き，従って一般化の能力が減るらしい，というようなことは直感的に予想されるが，こうしたことを明確に定式化するのは現在の課題である．

学習法自体の一般化については，すでにさまざまな一般化が提案されている．確率的降下法という本質を残しての一般化の方向としては，

- ユニットの一般化（準線形ユニット以外のユニットへの拡張）[32]
- ネットワークの形の一般化（回帰的な結合を含むネットワークへの拡張）[16][27][32]
- 学習の評価関数の一般化（二乗誤差以外の評価関数への拡張）

などが考えられている．

ニューラルネットの学習法としてはこの他にも，ボルツマンマシンの学習法[1]，競合学習[21]，連想記憶のための学習法[3][4][21]など多くの学習法が提案され，性質が研究されている．

4 おわりに

以上，現在までに提案されているニューラルネットによる情報処理および学習のモデルのごく一部をとりあげ，数理的な性質を簡単に紹介した．それぞれの詳細は参考文献を参照していただきたい．ここでとりあげたモデルにおいては，個々のユニット（神経細胞のモデル）は非常に簡単にモデル化されているが，それでも多くの未解決の問題がある．実際の神経回路網に触れながら実験を行っている研究者の間では，個々の神経細胞単位でもずっと複雑な情報処理が行われているという意見が多くあるが，現在までのところでは，そうした複雑なユニットのモデルを用いた情報処理のモデルの提案は少ない．

また，今回とりあげたモデルはいってみれば要素処理のモデルであり，これらのモデルのさらに上には，D.Marr のいう計算論的な問題，すなわち，脳が全体として（あるいは，ある程度のまとまりを持つ機能単位として）どのような情報処理をなぜ行っているのか？という問題がある[23]．たとえば，階層的なネットワークによるパターンの変換にしても，その原理的な能力からさらに踏み込んだ議論をするためには，脳が必要とする情報処理の性質をある程度一般的に規定できるような枠組みが必要であると感じられる．「脳はどのような原理に沿って何をしているのか？」という素朴な疑問に答えられるためにはより一層の研究が必要である．

参考文献

[1] Ackley,D.H., Hinton,G.E., and Sejnowski,T.J.: A learning algorithm for Boltzmann Machines, *Cognitive Science, Vol.9*, pp.147-169 (1985).

[2] 甘利俊一: 学習識別の理論, 電子通信学会誌, *Vol. 50*, pp.1272-1279 (1967).

[3] 甘利俊一: 神経回路網の数理－脳の情報処理様式－, 産業図書 (1978).

[4] Amari,S.: Neural theory of association and concept formation, *Biol. Cybern., Vol.26*, pp.175 -185 (1977).

[5] Amari,S. and Maginu,K.: Statistical neuro- dynamics of associative memory, *Neural Networks, Vol.1*, pp.63-73 (1988).

[6] Amit,D.J., Gutfreund,H., and Sompolinsky,H.: Storing infinite numbers of patterns in a spin- glass model of neural networks, *Phys. Rev. Letter, Vol.55*, pp.1530-1533 (1985).

[7] Cottrell,M.: Stability and attractivity in associative memory nettworks, *Biol Cybenetics, Vol.58*, pp.129-139 (1988).

[8] 船橋賢一: ニューラルネットワークの Capability について, 電子情報通信学会技術研究報告, *MBE88-52* (1988).

[9] Gallinari,P., Thiria,S., and Fogelmansoulie,F. : Multilayer perceptrons and data analysis, in *Proc. of IEEE ICNN88*, pp.I-391-399 (1988).

[10] Geman,S. and Geman,D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. PAMI, Vol.6*, pp.721-741 (1983).

[11] Haffner,P., Wibel,A., Shikano,T.: Fast back- propagation learning methods for neural networks in speech, 音響学会講演論文集 (1988/Oct.).

[12] Hinton,G.E., Sejnowski,T.J., and Ackley,D.H.: Boltzmann Machines: Constraint satisfaction networks that learn,*Tech. Rep. CMU-CS-84-119*, Carnegie-Mellon Univ. (1984).

[13] Hopfield,J.J.: Neural networks and physical systems with emergent collective computaional abilities,*Proceedings of the National Academy of Sciences USA, Vol.79*, pp.2254-2258 (1982).

[14] Hopfield,J.J., and Tank,D.W.: Neural computation of decisions in optimaization problems, *Bilogical Cybernetics, Vol.52*, pp.141-152 (1985).

[15] Irie,B. and Miyake,S.: Capabilities of three- layered Perceptrons, in *Proceedings of IEEE ICNN88*, pp.I-641 (1988).

[16] 河村嘉顕: ダイナミカルシステムとしてのニューロン回路の学習, 計測自動制御学会第 11 回 Dynamical System Symposium 予稿集 (1988).

//

[17] 木本 隆他: ニューラル・ネットワークの学習方式の検討, 昭和63年度人工知能学会全国大会予稿集, p.123-126 (1988).

[18] Kirkpatrick,S.,Gellat,C.D., and Vecci,M.P.: Optimization by simulated annealing,*Science, Vol.220*, pp.671-680 (1983).

[19] Kleene,S.C.: Representation of events in nerve nets and finite automata, in *Automata Studies* (Annals of Mathematical Studies no.34), Princeton (1956).

[20] Kohonen,T.: Correlation matrix memories, *IEEE Trans. Computers, Vol.21[4]*, pp.353-359 (1972).

[21] Kohonen,T.: *Self-organization and Associative Memory*, Springer-Verlag (1984).

[22] 倉田耕治: ボルツマンマシンの実数値化, 昭和63年電子情報通信学会春季全国大会予稿集, GD-5 (1988).

[23] Marr,D.: *Vision - A Computational Investi- gation into the Human Representation and Processing of Visual Information*, San Francisco: W.H.Freeman and Company (1982).

[24] McEliece,R.J., Posner,E.C., Rodemich,E.R., and Venkatesh,S.S.: The capacity of the Hopfield associative memory, *IEEE Trans. IT, Vol.33[4]*, pp.461-482 (1987).

[25] Minsky,M, and Papert,S.: *Perceptrons*, Cambridge, MA: MIT Press (1969).

[26] Nakano,K.: Associatron - a model of associative memory, *IEEE Trans. SMC, Vol.2*, pp.380-388 (1972).

[27] Pineda,F.J.: Generalization of back-propagation to recurrent neural networks, *Phisycs Review Letters, Vol.59*, pp.2229-2232 (1987).

[28] Ricotti,L.P., Ragazzini,S., and Martinelli,G.: Learning of word stress in a sub-optimal second order back-propagation neural network, in *Proceedings of IEEE ICNN88* (1988).

[29] Rosenblatt,F.: The perceptron: A probabilistic model for information storage and organization in the brain, *Psychological Rev., Vol.65[6]*, pp.386-408 (1958).

[30] Rosenblatt,F.: *Principles of Neurodynamics – Perceptrons and the Theory of Brain Mechanisms*, Washington,D.C: Spartan (1961).

[31] Rumelhart,D.E., Hinton,G.E., and Williams,R.J. : Learning representations by back-propagating errors, *Nature, Vol.323-9*, pp.533-536 (1986).

[32] Rumelhart,D.E., Hinton,G.E., and Williams,R.J. : Learning internal representaions by error propagation, in *Parallel Distributed Processing Volume 1*, J.L.McClelland, D.E.Rumelhart, and The PDP Research group, Cambridge, MA: MIT Press (1986).

[33] Szu,H. and Hartley,R.: Fast simulated annealing, *Physics Letters A, Vol.112*, pp.157-162 (1987).

[34] Takeda,M. and Goodman, J.W.: Neural networks for computation: number representation and programming complexity. *Applied Optics, Vol.25*, pp.3033-3064 (1986).

[35] 上坂, 尾関: 連想形記憶の二, 三の性質, 電子通信学会論文誌, *Vol.55-D*, pp.323-330 (1972).

[36] 上坂吉則: 2値変数の実数値関数から導かれるエネルギーを持つニューロン回路網の安定性について, 電子情報通信学会技術研究報告, *PRU88-6* (1988).

[37] Willshaw,D.J. and Longuet-Higgins,H.C.: Associative memory models, in *Machine Inteligence Vol.5*, pp.351-359, Meltzer,B. and Michie,D.(eds) Edinburgh Univ. Press (1969).